School of Computer Applications

Scheme & Syllabus

of

Master of Computer Applications (MCA)

2 Years Full time program

Specialization- Artificial Intelligence and Data Science

Academic Program

July 2024-26

JECRC University, Jaipur

Plot No. IS-2036 to IS-2039 Ramchandrapura Industrial Area Jaipur, Sitapura, Vidhani, Rajasthan 303905

Total Credits for the Batch 2024-27 = 80 Credits

SEMESTER WISE CREDITS				CREDITS
I	**II**	**III**	**IV**	**Total**
22	**19**	**23**	**16**	**80**

*SP-Specialization
*ID- Interdisciplinary
*F-Foundation
*GE- General Elective

AI-DS Specializations Subjects (MCA)							
Semester	**Code**	**Course Name**	**L (Hr.)**	**T (Hr.)**	**P (Hr.)**	**Credits**	**Type**
I	**MCA520A/B**	**Introduction to Artificial Intelligence Machine Learning and Data Science**	**4**	**0**	**2**	**5**	**SP**
II	**MCA521A/B**	**Python for Artificial Intelligence and Data Science**	**4**	**0**	**2**	**5**	**SP**
II	**MCA525A/B**	**Advance Data Science**	**4**	**0**	**2**	**5**	**SP**
III	**MCA522A**	**Advance Linear Algebra Probability and Statistics for Machine Learning**	**4**	**0**	**0**	**4**	**SP**
III	**MCA524A/B**	**Deep Learning**	**4**	**0**	**2**	**5**	**SP**
Total			**20**	**0**	**6**	**24**	

Semester – I								
S. No.	**Code**	**Course Name**	**L (Hr.)**	**T (Hr.)**	**P (Hr.)**	**Credits**	**Contact Hours**	**Type**
1	MCA125B	Competitive Programming (Using C++)	4	0	0	4	4	CORE
2	**MCA520A**	**Introduction to Artificial Intelligence Machine Learning and Data Science**	**4**	**0**	**0**	**4**	4	**SP**
3	MCA121B	Advance Data Structures and Algorithms	4	0	0	4	**4**	CORE
4	MCA201 A	Advance Operating System	4	0	0	4	4	CORE
5	MCA129B	Competitive Programming (Using C++) Lab	0	0	2	1	2	CORE
6	MCA203A	Advance Data Structures Lab	0	0	2	1	**2**	CORE
7	**MCA520B**	**Introduction to Artificial Intelligence Machine Learning and Data Science Lab**	**0**	**0**	**2**	**1**	2	**SP**
8	DEN004A	Professional Communication Skills	2	0	2	3	5	ID
Total			**18**	**0**	**8**	**22**	**27**	

Semester – II								
S. No.	Code	Course Name	L (Hr.)	T (Hr.)	P (Hr.)	Credits	Contact Hours	Type
1	**MCA521A**	**Python for Artificial Intelligence and Data Science**	**4**	**0**	**0**	**4**	**4**	**SP**
2	MCA118A	Advance Database Management System	4	0	0	4	4	CORE
3	MCA130A	Advance Java	4	0	0	4	**4**	CORE
5	**MCA525A**	**Advance Data Science**	**4**	**0**	**0**	**4**	4	**SP**
6	**MCA521B**	**Python for Artificial Intelligence and Data Science Lab**	**0**	**0**	**2**	**1**	**2**	**SP**
7	MCA124C	Project Lab in Advance Database Management Systems	0	0	2	1	2	CORE
8	**MCA525B**	**Advance Data Science Lab**	**0**	**0**	**2**	**1**	2	**SP**
9	MCA202A	Advance Java Lab	0	0	2	1	2	CORE
Total			**16**	**0**	**08**	**20**	**24**	

Semester – III								
S. No.	**Code**	**Course Name**	**L (Hr.)**	**T (Hr.)**	**P (Hr.)**	**Credits**	**Contact Hours**	**Type**
1	MCA312A	Web Technology	4	0	0	4	4	CORE
2	**MCA524A**	**Deep Learning**	**4**	**0**	**0**	**4**	**4**	**SP**
3	**MCA522A**	**Advance Linear Algebra Probability and Statistics for Machine Learning**	**4**	**0**	**0**	**4**	**4**	**SP**
4	MCA313A	Web Technology Lab	0	0	2	1	2	CORE
5	**MCA524B**	**Deep Learning Lab**	**0**	**0**	**2**	**1**	**2**	**SP**
6	MCA316A	Project	0	0	4	2	4	CORE
7		Department Elective-III	4	0	0	4	4	E
8		Technical Research Paper Writing	2	0	0	2	2	CORE
Total			**18**	**0**	**08**	**22**	**26**	

Semester – IV					
Course Code	**Course Name**	**L (Hr.)**	**P (Hr.)**	**Credits**	**Type**
MCA175A	**Industrial Training/Internship**	**0**	**0**	**16**	**CORE**

SYLLABUS

SEMESTER - I

MCA520A

Introduction to Artificial Intelligence Machine Learning and Data Science

Unit I	**Introduction to Artificial Intelligence** Definition and scope of Artificial Intelligence, Historical background and milestones in AI development, Various branches of AI: symbolic AI, statistical AI, etc., Applications of AI in different fields like healthcare, finance, gaming, etc., Ethical considerations and societal impact of AI
Unit II	**Fundamentals of Machine Learning** Introduction to Machine Learning and its importance, Types of Machine Learning: Supervised, Unsupervised, and Reinforcement Learning, Basic concepts: features, labels, training data, etc., Popular Machine Learning algorithms: Linear Regression, Logistic Regression, Decision Trees, k-Nearest Neighbors, etc., Evaluation metrics for Machine Learning models: accuracy, precision, recall, F1-score, etc.
Unit III	**Machine Learning Techniques** Data preprocessing techniques: handling missing data, feature scaling, feature encoding, etc., Model selection and hyperparameter tuning, Cross-validation techniques, Ensemble methods: Bagging, Boosting, Random Forests, etc., Introduction to deep learning and neural networks
Unit IV	**Introduction to Data Science** What is Data Science and why it is important?, Role of Data Scientist and skills required, Data acquisition: sources of data, data formats, data cleaning, etc., Exploratory Data Analysis (EDA): statistical analysis, data visualization techniques, Introduction to libraries/tools: NumPy, Pandas, Matplotlib, Seaborn, etc
Unit V	**Advanced Topics and Applications** Advanced Machine Learning techniques: Support Vector Machines (SVM), Neural Networks, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), etc., Natural Language Processing (NLP) and its applications, Introduction to Big Data technologies: Hadoop, Spark, etc., Case studies and real-world applications in various domains, Future trends and career prospects in AI, ML, and Data Science

Reference Books:

1. "Artificial Intelligence: A Modern Approach" by Stuart Russell and Peter Norvig
2. "Introduction to Machine Learning" by Ethem Alpaydin

3. "Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett

SEMESTER I

MCA522A

Advance Linear Algebra Probability and Statistics for Machine Learning

Unit I	**Advanced Linear Algebra** Vector spaces and subspaces, Matrix factorization techniques: LU decomposition, QR decomposition, Singular Value Decomposition (SVD), Eigenvalues and eigenvectors: properties and applications, Positive definite matrices and their properties, Applications of linear algebra in machine learning: principal component analysis (PCA), linear regression, support vector machines
Unit II	**Probability Theory for Machine Learning** Probability axioms and rules, Random variables and probability distributions: discrete and continuous distributions, Joint, marginal, and conditional probability, Expectation, variance, covariance, and correlation, Bayes' theorem and its applications in machine learning
Unit III	**Statistical Inference** Point estimation: maximum likelihood estimation (MLE), method of moments, Interval estimation: confidence intervals, Hypothesis testing: null and alternative hypotheses, type I and type II errors, p-values, Parametric and non-parametric tests: t-tests, chi-square tests, Wilcoxon rank-sum test, Applications of statistical inference in machine learning: hypothesis testing for model comparison, parameter estimation
Unit IV	**Advanced Topics in Probability and Statistics** Bayesian inference: Bayesian parameter estimation, Bayesian model comparison, Markov chains and Markov processes, Hidden Markov models (HMMs) and their applications, Time series analysis: autoregressive models, moving average models, ARIMA models, Monte Carlo methods: Markov Chain Monte Carlo (MCMC), Metropolis-Hastings algorithm
Unit V	**Applications of Probability and Statistics in Machine Learning** Probabilistic graphical models: Bayesian networks, Markov random fields, Gaussian processes and their applications, Ensemble methods: bagging, boosting, stacking, Cross-validation techniques: k-fold cross-validation, leave-one-out cross-validation, Evaluation metrics for machine learning models: accuracy, precision, recall, F1-score, ROC curves, AUC-ROC

Reference Books:

1) "Linear Algebra and Its Applications" by Gilbert Strang
2) "Probability and Statistics for Computer Scientists" by Michael Baron
3) "Pattern Recognition and Machine Learning" by Christopher M. Bishop

SEMESTER II

MCA521A

Python for Artificial Intelligence and Data Science

Unit I	**Introduction to Python Programming** Introduction to Python: history, features, and advantages, Setting up Python development environment (IDEs, Jupyter Notebooks, etc.), Basic Python syntax: variables, data types, operators, control structures (if, else, loops), Functions and modules in Python, Handling exceptions in Python programs
Unit II	**Python Libraries for Data Handling and Manipulation** Introduction to NumPy for numerical computing with Python, Working with arrays and matrices in NumPy, Introduction to Pandas for data manipulation and analysis, Data structures in Pandas: Series, DataFrame, Index, Data cleaning and preprocessing techniques using Pandas
Unit III	**Advanced Data Visualization with Python** Interactive visualization with Plotly and Bokeh, Customizing visualizations with Matplotlib and Seaborn: advanced plotting techniques, themes, styles, 3D visualization with Matplotlib and Plotly, Dashboards and interactive applications using Dash, Introduction to geospatial data visualization with GeoPandas and Folium
Unit IV	**Advanced Machine Learning Techniques** Ensemble learning methods: Random Forest, Gradient Boosting, AdaBoost, Kernel methods: Support Vector Machines (SVM), Kernel Ridge Regression, Unsupervised learning techniques: Clustering (K-Means, DBSCAN), Dimensionality Reduction (PCA, t-SNE), Semi-supervised and self-supervised learning, Anomaly detection algorithms
Unit V	**Deep Learning with Python** Advanced topics in deep learning: advanced architectures (Capsule Networks, Transformer Networks), attention mechanisms, adversarial training, Advanced techniques for training deep neural networks: transfer learning, curriculum learning, learning rate schedules, Model optimization techniques: model compression, quantization, pruning, Introduction to distributed deep learning with TensorFlow and PyTorch, Advanced applications of deep learning: natural language processing, computer vision, reinforcement learning

Reference Books:

1) "Python for Data Analysis" by Wes McKinney
2) "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
3) "Pattern Recognition and Machine Learning" by Christopher M. Bishop
4) "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

SEMESTER II

MCA525A

Advance Data Science

Unit I	**Advanced Data Manipulation and Analysis** Advanced data manipulation with Pandas: groupby, pivot tables, time series data, Working with large datasets: memory optimization techniques, Advanced data cleaning techniques: handling missing data, outlier detection and removal, Merging and joining datasets, Introduction to data serialization formats: JSON, XML, YAML
Unit II	**Advanced Data Visualization** Interactive visualization with Plotly and Bokeh, Customizing visualizations with Matplotlib and Seaborn: advanced plotting techniques, themes, styles, 3D visualization with Matplotlib and Plotly, Dashboards and interactive applications using Dash, Introduction to geospatial data visualization with GeoPandas and Folium
Unit III	**Machine Learning Techniques** Advanced machine learning techniques: ensemble methods (Random Forests, Gradient Boosting), feature engineering, model stacking, Hyperparameter optimization techniques: grid search, randomized search, Bayesian optimization, Model interpretation and explainability techniques, Introduction to imbalanced data handling techniques, Machine learning pipelines and workflows
Unit IV	**Deep Learning Fundamentals** Introduction to neural networks and deep learning, Convolutional Neural Networks (CNNs) for image data, Recurrent Neural Networks (RNNs) for sequential data, Transfer learning and pre-trained models, Advanced topics: Generative Adversarial Networks (GANs), Reinforcement Learning

Unit V	**Big Data Analytics and Scalable Data Science** Distributed computing frameworks: Apache Spark, Hadoop, Scalable machine learning algorithms for big data, Streaming data processing and real-time analytics, Introduction to cloud-based data analytics platforms: AWS, Google Cloud, Microsoft Azure, Ethical considerations and challenges in Data Science

Reference Books:

1) "Python for Data Analysis" by Wes McKinney
2) "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
3) "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

SEMESTER III

MCA524A

Deep Learning

Unit I	**Introduction to Deep Learning** Definition and importance of Deep Learning, Historical overview and milestones in Deep Learning, Comparison with traditional machine learning approaches, Basic concepts: neural networks, activation functions, layers, etc., Applications of Deep Learning in various domains
Unit II	**Fundamentals of Neural Networks** Perceptrons and the McCulloch-Pitts model, Feedforward neural networks: architecture, forward propagation, backpropagation, Activation functions: sigmoid, tanh, ReLU, etc., Loss functions and optimization techniques: gradient descent, stochastic gradient descent, etc., Regularization techniques: dropout, L2 regularization, etc.
Unit III	**Convolutional Neural Networks (CNNs)** Introduction to CNNs and their applications in computer vision, Convolutional layers: filters, feature maps, receptive fields, Pooling layers: max pooling, average pooling, CNN architectures: LeNet, AlexNet, VGG, ResNet, etc., Transfer learning and fine-tuning pre-trained CNNs
Unit IV	**Recurrent Neural Networks (RNNs)** Introduction to RNNs and their applications in sequential data analysis, Basic RNN architecture: recurrent connections, hidden states, time-series data, Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), Applications of RNNs: natural language processing, time series prediction, etc., Challenges and limitations of RNNs

Unit V	**Advanced Topics in Deep Learning** Generative Adversarial Networks (GANs) and their applications in generative modeling, Autoencoders and dimensionality reduction techniques, Attention mechanisms in Deep Learning, Reinforcement Learning and its integration with Deep Learning, Future trends and emerging technologies in Deep Learning

Reference Books:

1) "Deep Learning for Computer Vision" by Rajalingappaa Shanmugamani
2) "Neural Networks and Deep Learning: A Textbook" by Charu C. Aggarwal
3) "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

SEMESTER III

MCA523A

Data Visualization Tools and Techniques

Unit I	**Introduction to Data Visualization** Definition and importance of data visualization, Principles of effective data visualization, Types of data visualization: charts, graphs, maps, etc., Data visualization process: data acquisition, exploration, analysis, visualization, interpretation, Ethical considerations and best practices in data visualization
Unit II	**Basic Data Visualization Techniques** Introduction to basic visualization tools: Excel, Google Sheets, Tableau Public, etc., Creating and customizing simple charts: bar charts, line charts, pie charts, scatter plots, etc., Adding labels, titles, legends, and annotations to visualizations, Data filtering and sorting techniques for effective visualization, Exporting and sharing visualizations
Unit III	**Advanced Data Visualization Techniques** Introduction to advanced visualization libraries: Matplotlib, Seaborn, Plotly, etc., Creating complex and interactive visualizations using Python, Heatmaps, box plots, violin plots, histograms, etc., Geographic visualization: maps, choropleth maps, geospatial data visualization, Dashboard creation and customization
Unit IV	**Data Visualization Best Practices** Choosing the right visualization for different types of data and analysis goals, Design principles for effective visualization: color theory, layout, typography, etc., Data storytelling: conveying insights and narratives through visualizations, Accessibility and inclusivity considerations in data visualization, Critique and evaluation of data visualizations

Unit V	**Data Visualization Projects and Applications** Hands-on projects to apply data visualization techniques to real-world datasets, Exploratory data analysis (EDA) through visualization, Case studies and examples of data visualization in various domains: business, healthcare, finance, etc., Future trends and emerging technologies in data visualization, Career opportunities and pathways in data visualization and related fields

Reference Books:

1) "Data Visualization: A Practical Introduction" by Kieran Healy
2) "Python Data Visualization: An Introduction to Data Visualization in Python with Matplotlib, Seaborn, and Plotly" by Adel Osmani
3) "The Visual Display of Quantitative Information" by Edward Tufte

Practical

Subject	**Introduction to Artificial Intelligence Machine Learning and Data Science Lab**
Code	**MCA520B**
Credit	**1**

Course Outcomes:

Students explore supervised and unsupervised learning techniques, and apply advanced concepts to real-world scenarios. By mastering data analysis and visualization, students gain the skills needed to excel in diverse industries.

Exercises—

1) Load a dataset (e.g., Iris dataset, Titanic dataset) into Python using Pandas. Explore the dataset by:
    - Displaying the first few rows.
    - Checking for missing values.
    - Calculating summary statistics.
    - Visualizing distributions of numerical variables.
2) Perform data preprocessing tasks such as:

- Handling missing values (e.g., imputation, deletion).
- Encoding categorical variables (e.g., one-hot encoding, label encoding).
- Scaling numerical features (e.g., standardization, normalization).

3) Implement a supervised learning algorithm (e.g., decision tree classifier, logistic regression) using Scikit-learn. Train the model on a training set and evaluate its performance on a test set using appropriate metrics (e.g., accuracy, precision, recall, F1-score).
4) Experiment with different hyperparameters of the supervised learning model (e.g., max_depth for decision trees, C for logistic regression) and observe how they affect model performance.
5) Apply an unsupervised learning algorithm (e.g., K-means clustering, hierarchical clustering) to a dataset. Explore the resulting clusters and interpret the findings.
6) Use dimensionality reduction techniques (e.g., Principal Component Analysis, t-SNE) to visualize high-dimensional data in two or three dimensions. Discuss the insights gained from the visualization.
7) Split a dataset into training and testing sets using cross-validation techniques (e.g., k-fold cross-validation, stratified cross-validation). Train a machine learning model on multiple folds and compute the average performance metrics.
8) Perform model selection by comparing the performance of multiple algorithms (e.g., decision tree, random forest, support vector machine) on a given dataset. Choose the best-performing model based on evaluation metrics.
9) Explore ensemble learning techniques (e.g., bagging, boosting) by implementing ensemble models such as Random Forest or Gradient Boosting Machines. Compare the performance of ensemble models with individual base models.
10) Implement a neural network model using TensorFlow or PyTorch. Design the architecture of the neural network (e.g., number of layers, activation functions) and train the model on a dataset. Evaluate the performance of the neural network on a test set.
11) Choose a real-world dataset related to a specific domain (e.g., healthcare, finance, retail). Apply appropriate machine learning techniques to solve a relevant problem, such as predicting customer churn, diagnosing diseases, or forecasting stock prices.
12) Present the results of your analysis in a clear and interpretable manner, using data visualization technique ues (e.g., plots, charts) to communicate insights to stakeholders.

Subject	**Python for Artificial Intelligence and Data Science Lab**
Code	**MCA521B**
Credit	**1**

Course Outcomes:

Students learn model evaluation, optimization, and ethical considerations, fostering critical thinking and effective communication for real-world applications. By mastering these concepts, students are prepared to tackle complex data challenges and contribute responsibly to the field of Artificial Intelligence and Data Science.

Exercises—

1) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
2) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
3) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
4) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
5) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
6) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
7) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
8) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
9) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.

10) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
11) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
12) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.
13) Load a dataset using Pandas and perform basic data exploration tasks such as checking for missing values and visualizing data distributions.
14) Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features.

Subject	**Advance Data Science Lab**
Code	**MCA525B**
Credit	**1**

Course Outcomes: Students with proficient skills in cutting-edge data analysis techniques, including feature engineering, time series analysis, and deep learning. Through practical exercises, students develop expertise in deploying machine learning models, interpreting complex data patterns, and solving real-world problems across diverse domains. By mastering these advanced techniques, students are prepared to excel in industry roles requiring advanced data analysis and modeling capabilities.

Exercises—

1) Implement feature engineering techniques such as polynomial features, interaction terms, and dimensionality reduction.
2) Use feature selection methods like Recursive Feature Elimination (RFE) or feature importance from tree-based models to identify the most relevant features for predictive modeling**.**
3) Explore time series decomposition techniques to identify trends, seasonality, and noise in temporal data.
4) Implement forecasting models such as ARIMA, SARIMA, or Prophet to predict future values in time series data.
5) Preprocess text data by tokenization, stemming, and removing stopwords.

6) Build sentiment analysis or text classification models using techniques like Bag-of-Words, TF-IDF, or word embeddings (Word2Vec, GloVe).
7) Apply clustering algorithms (e.g., K-means, DBSCAN) to identify natural groupings in data.
8) Implement anomaly detection techniques (e.g., Isolation Forest, One-Class SVM) to detect outliers or anomalies in datasets.
9) Work with large-scale datasets using distributed computing frameworks like Apache Spark.
10) Implement parallelized algorithms for data processing, analysis, and modeling on data platforms.

Subject	**Deep Learning Lab**
Code	**MCA524B**
Credit	**1**

Course Outcomes:

Exercises—

1) Implement a feedforward neural network using a deep learning framework like TensorFlow or PyTorch.
2) Train the neural network on a dataset (e.g., MNIST handwritten digits) for a classification task.
3) Experiment with different network architectures, activation functions, and optimization algorithms**.**
4) Implement a CNN architecture for image classification using TensorFlow or PyTorch.
5) Train the CNN on a dataset such as CIFAR-10 or ImageNet.
6) Fine-tune hyperparameters such as kernel size, number of filters, and pooling layers to improve model performance.
7) Implement an RNN architecture for sequence prediction tasks like text generation or sentiment analysis.
8) Train the RNN on a text dataset (e.g., movie reviews or news articles).
9) Experiment with different types of RNN cells (e.g., LSTM, GRU) and varying sequence lengths.

10) Fine-tune a pre-trained deep learning model (e.g., VGG16, ResNet) on a new dataset with limited labeled samples.
11) Evaluate the performance of the transferred model on the new dataset and compare it with training from scratch.
12) Perform hyperparameter tuning using techniques like grid search or random search to optimize model performance.
13) Tune hyperparameters such as learning rate, batch size, and dropout rate for a deep learning model on a specific task.
14) Deploy a trained deep learning model in a production environment using frameworks like TensorFlow Serving or ONNX Runtime.
15) Optimize the model for inference speed and memory footprint using techniques like quantization or model pruning.